

Bab II

Tinjauan Pustaka

Bab ini menjelaskan teori yang digunakan dalam mendukung penelitian tugas akhir ini. Teori mengenai status gizi balita, machine learning dan klasifikasi Naive Bayes.

2.1 Penelitian Terkait

Sebelumnya pada tahun 2015. Apriliyani Fitri Cahyanti menulis penelitian dengan tema “ Penentuan Model Terbaik dalam Menentukan Status Gizi Balita dengan Mempertimbangkan Independensi Parameter”[7]. Independensi parameter sendiri ditentukan berdasarkan nilai kolerasi atau kemiripan parameter yang digunakan dalam menentukan status gizi balita. Penelitian ini membahas mengenai uji korelasi antar parameter dalam studi kasus penilaian status gizi menggunakan metode cosine similarity. Penggunaan metode *Cosine Similarity* untuk mengetahui similaritas antar parameter, dan kolerasi antar parameter tersebut, sehingga dapat dijadikan sebagai penentuan independensi antar parameter. Dari hasil uji korelasi dengan *Cosine Similarity*, parameter dengan nilai similarity terbesar memiliki pengaruh terhadap hasil penilaian status gizi balita. Penilaian status gizi sendiri berdasarkan kategori pengukuran BMI(*Body Mass Index*) dimana parameter yang digunakan yaitu berat badan dan umur.. Hasil pengolahan dan pengujian yang dilakukan diperoleh berupa model terbaik dari penilaian status gizi balita dilihat dari akurasi, kesederhanaan, waktu dan akuisisi data. Ditunjukan pada model yang terdiri dari parameter berat badan, BMI (*Body Mass Index*), dan umur dengan akurasi sebesar 61%. Penelitian ini menunjukan bahwa penerapan kolerasi antar parameter dengan metode *Cosine Similarity* dapat digunakan dalam pembentukan model terhadap penilaian status gizi balita.

Pada tahun 2015 Abdul Wafi [8] membuat sistem perangkat lunak dengan tema “Penentuan Status Gizi dan Tinggi Balita Menggunakan Metode *Naïve Bayes Classifier* Berbasis Web”. Sistem yang digunakan dalam membangun aplikasi website berupa pemrograman PHP dan *MySql* sebagai database, sedangkan untuk sistem klasifikasinya sendiri menggunakan *Naive Bayes*

Classifier. Metode *Naive Bayes Classifier* merupakan metode machine learning yang digunakan untuk mengklasifikasikan suatu permasalahan dengan pembentukan model dimana menghasilkan nilai akurasi yang dapat digunakan untuk mengambil sebuah keputusan. Variable yang digunakan untuk penentuan status gizi yaitu jenis umur, berat badan dan tinggi badan. Pada penelitian ini penentuan status gizi balita diklasifikasikan berdasarkan indeks berat badan terhadap umur dan tinggi badan terhadap umur. Hasil klasifikasi penentuan status gizi yang di hasilkan berdasarkan indeks berat badan terhadap umur antara lain gizi buruk, gizi normal, gizi lebih. Sedangkan hasil klasifikasi berdasarkan indeks tinggi badan terhadap umur antara lain sangat pendek, pendek dan normal. Tahap implementasi sistem pada aplikasi website dilakukan setelah dilakukan klasifikasi pada data, untuk mengetahui keberhasilan implementasi sistem maka dilakukan pengujian nilai akurasi. Pengujian dilakukan pada 150 data dengan membagi dataset menjadi 80% data sebagai data training, dan 20% data sebagai data testing. Nilai akurasi dari pengujian didapatkan sebesar 41,33% pada indeks tinggi badan terhadap umur, dan 47,99% pada indeks berat badan terhadap umur.

Nur Rachmaliany, pada tahun 2017 [9], melakukan pengembangan aplikasi website untuk penentuan nutrisi pada anak dengan metode *Fuzzy C-Means* berdasarkan produk kemasan. *Fuzzy C-Means* merupakan teknik machine learning dengan sistem clustering data yang mana keberadaan tiap titik data dalam suatu cluster ditentukan oleh derajat keanggotaan. Penggunaan metode FCM (*Fuzzy C-Means*) dalam sistem ini membantu *cluster* atau pengelompokan status gizi, dengan mempertimbangkan beberapa atribut seperti berat badan, tinggi badan dan kandungan nutrisi dalam makanan kemasan. Kandungan nutrisi dalam kemasan digunakan untuk menganalisis kandungan nutrisi yang dibutuhkan anak serta sebagai bahan pertimbangan dalam penentuan status gizi. Selain itu ditambahkan variabel lain berupa tiga buah kormobid (diare, edema dan pneumonia) sebagai atribut tambahan. Setelah menentukan pengembangan sistem yang dilakukan dengan algoritma *Fuzzy C-Means*, selanjutnya pengaplikasian sistem dengan menggunakan bahasa pemrograman php, sebagai program, kemudian data ditampilkan ke dalam sebuah web server dengan melakukan komputerisasi terhadap seluruh data informasi yang didapatkan dari penelitian sehingga

dihasilkan suatu sistem penentuan nutrisi secara tepat dan akurat. Hasil dari sistem clustering berupa status gizi kurang, gizi baik dan gizi lebih. Untuk mengoptimalkan sistem bekerja dengan baik, maka pengujian dilakukan dengan *User Acceptance Test* atau pengujian terhadap kinerja software, *Usability Testing* untuk mengetahui kemudahan penggunaan aplikasi, dan validasi *Fuzzy C-Means*. Dari hasil pengujian didapatkan nilai *User Acceptance Test* sebesar 89%, *Usability Testing* 74% dan penerapan *Fuzzy C-Means* sebesar 61%. Hal ini menunjukkan bahwa aplikasi website dapat bekerja dengan baik namun dalam penerapan metode *Fuzzy C-Means* dinilai belum cukup baik dalam clustering status gizi balita berdasarkan kandungan nutrisi makanan kemasan.

2.2 Gizi

Gizi berasal dari kata bahasa Arab "Ghidza" yang berarti makanan yang berkaitan dengan makanan dan berkaitan pula dengan tubuh manusia. Kata gizi selain berkaitan dengan kesehatan juga berkaitan dengan potensi ekonomi seseorang, yaitu berhubungan dengan perkembangan otak, kemampuan belajar dan produktifitas kerja[10]. Menurut Sunita Almatsie Gizi adalah ilmu yang mempelajari segala sesuatu tentang makanan dalam hubungannya dengan kesehatan optimal. Zat gizi menghasilkan ikatan kimia yang diperlukan tubuh untuk melakukan fungsinya, yaitu menghasilkan energi, membangun dan memelihara jaringan serta mengatur proses-proses kehidupan[10].

Sedangkan menurut WHO gizi merupakan pilar utama dari kesehatan dan kesejahteraan sepanjang siklus kehidupan, sejak janin dalam kandungan, bayi, balita, anak, remaja dan usia lanjut[11]. Makanan yang memenuhi syarat gizi merupakan kebutuhan utama untuk pertahanan hidup, pertumbuhan fisik, perkembangan mental, kesehatan dan kesejahteraan.

Untuk memenuhi keseimbangan zat gizi dalam tubuh asupan makanan harus mengandung semua jenis zat gizi, antara lain : karbohidrat, lemak, protein, vitamin, mineral dan air. Karbohidrat merupakan jenis zat gizi yang memegang peranan penting dalam kehidupan karena merupakan sumber energi utama. Selain karbohidrat, lemak juga penghasil energi yang memberikan kontribusi terhadap

tubuh. Sebagai zat yang membangun sel-sel tubuh, protein merupakan zat gizi penting lain yang harus terdapat dalam konsumsi makanan sehari-hari, di samping vitamin dan mineral yang berperan dalam zat pengatur metabolisme.

2.2.1 Status Gizi

Status gizi merupakan keadaan yang diakibatkan oleh keseimbangan antara asupan zat gizi dari makanan dengan kebutuhan zat gizi yang diperlukan untuk metabolisme tubuh [10, p. 10]. Setiap individu membutuhkan asupan gizi yang berbeda antar individu. Hal ini tergantung pada usia, jenis kelamin dan aktivitas tubuh dalam sehari. Dalam penentuan status gizi pada balita, status gizi digunakan untuk mengetahui kesehatan anak. Secara umum berdasarkan peraturan kementerian kesehatan no.2 tahun 2019 tentang standar antropometri anak, status gizi pada balita dapat ditentukan berdasarkan 4 indeks, antara lain berat badan menurut umur (BB/U), tinggi badan menurut umur (TB/U), berat badan menurut umur (BB/TB) dan indeks massa tubuh menurut umur (IMT/U) [12].

a. Berdasarkan Indikator BB/U

Indikator berat badan menggambarkan berat badan relatif dibandingkan dengan umur anak. Berat badan menggambarkan massa tubuh, dimana massa tubuh sangat sensitif terhadap perubahan-perubahan yang mendadak [13]. Misalnya karena penyakit, infeksi, penurunan nafsu makan atau penurunan jumlah zat gizi yang dikonsumsi. Selain itu parameter berat badan merupakan parameter yang paling sederhana, dan mudah diukur. Indikator berat badan menurut umur BB/U juga merupakan parameter yang sangat labil. Dalam keadaan normal, dimana kesehatan baik dan seimbang, maka berat badan mengikuti pertambahan umur. Sebaliknya jika keadaan abnormal, maka kemungkinan berat badan berkembang cepat atau lebih lambat. Dikarenakan indikator berat badan ini cenderung labil dan cepat untuk berubah. Maka penentuan status gizi berdasarkan berat badan lebih menggambarkan status gizi balita saat ini. Data baku berat badan menurut umur pada balita berdasarkan Kementerian Kesehatan [12, p. 16] dapat dilihat pada tabel 2.1.

Tabel 2.1 Data Baku Kementerian Kesehatan untuk Berat Badan menurut Umur (BB/U)

Umur (bulan)	Berat Badan (KG)						
	-3 SD	-2 SD	-1SD	Median	+1SD	+2SD	+3SD
0	2.1	2.5	2.9	3.3	3.9	4.4	5.0
1	2.9	3.4	3.9	4.5	5.1	5.8	6.6
2	3.8	4.3	4.9	5.6	6.3	7.1	8.0
3	4.4	5.0	5.7	6.4	7.2	8.0	9.0
4	4.9	5.6	6.2	7.0	7.8	8.7	9.7
5	5.3	6.0	6.7	7.5	8.4	9.3	10.4
6	5.7	6.4	7.1	7.9	8.8	9.8	10.4
7	5.9	6.7	7.4	8.3	9.2	10.3	11.4
8	6.2	6.9	7.7	8.6	9.6	10.7	11.9
9	6.4	7.1	8.0	8.9	9.9	11.0	12.3

Dari data baku acuan penentuan status gizi berdasarkan berat badan menurut umur, dapat digunakan untuk menilai status gizi balita dengan kondisi berat badan kurang, atau sangat kurang. Tetapi tidak dapat digunakan untuk mengklasifikasikan balita dengan kondisi gemuk atau sangat gemuk. Maka dari itu perlu dilakukan pengukuran tinggi badan. Selain itu harus melihat indikator BB/TB atau IMT/U. Tabel 2.1 merupakan contoh sebagian data baku Kementerian Kesehatan yang digunakan sebagai acuan dalam menentukan status gizi balita berdasarkan berat badan menurut umur.

b. Berdasarkan Indikator Tinggi Badan menurut Umur (TB/U)

Indikator tinggi badan menurut umur menggambarkan pertumbuhan panjang atau tinggi badan anak berdasarkan umurnya. Tinggi badan relatif kurang sensitif terhadap masalah kekurangan gizi dalam jangka waktu yang pendek. Sehingga indikator tinggi badan dapat mengidentifikasi balita dengan kondisi tubuh pendek (stunted) atau sangat pendek, yang disebabkan oleh kekurangan gizi dalam waktu yang lama[13]. Selain itu indikator tinggi badan menggambarkan status gizi masa lampau, status sosial dan adanya masalah gizi

kronis. Standar baku penentuan status gizi berdasarkan indikator tinggi badan dapat dilihat berdasarkan data baku kementerian kesehatan seperti pada tabel 2.2.

Tabel 2.2 Data Baku Kementerian Kesehatan untuk Tinggi Badan menurut Umur (TB/U)

Umur (bulan)	Tinggi Badan(TB)						
	-3 SD	-2 SD	-1SD	Median	+1SD	+2SD	+3SD
0	44.2	46.1	48.0	49.9	51.8	53.7	55.6
1	48.9	50.8	52.8	54.7	56.7	58.6	60.6
2	52.4	54.4	54.4	58.4	60.4	62.4	64.4
3	55.3	57.3	56.4	61.4	63.5	65.5	67.6
4	57.6	59.7	59.7	63.9	66.0	68.0	70.1
5	59.6	61.7	61.7	65.9	68.0	70.1	72.2
6	61.2	63.3	63.3	67.6	69.8	71.9	74.0
7	62.7	64.8	67.0	69.2	71.3	73.5	75.7
8	64.0	66.2	68.4	70.6	72.8	75.0	77.2
9	65.2	67.5	69.7	72.0	74.2	76.5	78.7

Tabel 2.1 merupakan contoh sebagian data baku Kementerian Kesehatan yang digunakan sebagai acuan dalam menentukan status gizi balita berdasarkan tinggi badan menurut umur.

c. Berdasarkan Indikator Berat Badan menurut Tinggi Badan (BB/TB)

Indikator berat badan menurut tinggi badan ini menggambarkan apakah berat badan anak sesuai terhadap pertumbuhan panjang/tinggi badan[13]. Indikator ini merupakan indikator terbaik untuk menilai status gizi saat ini dengan lebih spesifik. Terutama bila data umur sulit untuk diperoleh, selain itu indikator BB/TB menggambarkan ada tidaknya suatu masalah gizi akut, dan dapat membantu menentukan apakah berat badan anak masih dalam kisaran yang sesuai dengan tinggi badannya. Status gizi yang dihasilkan berdasarkan indikator BB/TB antara lain kondisi balita dengan gizi kurang, gizi buruk, serta balita dengan resiko gizi lebih. Kondisi gizi buruk biasanya disebabkan oleh

adanya faktor lain. Contohnya penyakit atau kekurangan asupan gizi yang baru saja terjadi maupun yang telah lama terjadi/kronis. Standar baku penentuan status gizi berdasarkan indikator berat badan menurut tinggi badan (BB/TB) dapat dilihat berdasarkan data baku kementerian kesehatan[12] seperti pada tabel 2.3.

Tabel 2.3 Data Baku Kementerian Kesehatan untuk Berat Badan menurut Tinggi Badan (BB/TB)

Tinggi Badan	Berat Badan(Kg)						
	-3 SD	-2 SD	-1SD	Median	+1SD	+2SD	+3SD
45.0	1.9	2.0	2.2	2.4	2.7	3.0	3.3
45.5	1.9	2.1	2.3	2.5	2.8	3.1	3.4
46.0	2.0	2.2	2.4	2.6	2.9	3.1	3.5
46.5	2.1	2.3	2.5	2.7	3.0	3.2	3.6
47.0	2.1	2.3	2.5	2.8	3.0	3.3	3.7
47.5	2.2	2.4	2.6	2.9	3.1	3.4	3.8
48.0	2.3	2.5	2.7	2.9	3.2	3.6	3.9
48.5	2.3	2.6	2.8	3.0	3.3	3.7	4.0
49.0	2.4	2.6	2.9	3.1	3.4	3.8	4.2
49.5	2.5	2.7	3.0	3.2	3.5	3.9	4.3

Tabel 2.3 merupakan contoh sebagian data baku Kementerian Kesehatan yang digunakan sebagai acuan dalam menentukan status gizi balita berdasarkan tinggi badan menurut umur. Untuk lebih lengkapnya dapat dilihat pada lampiran 1.

2.2.2 Kategori dan Ambang Batas Status Gizi Anak

Dari standar baku data berat badan maupun tinggi badan berdasarkan BB/U, TB/U dan BB/U. Maka penentuan status gizi ditentukan dengan ambang batas yang juga diatur berdasarkan peraturan Kementerian Kesehatan no 2 tahun 2020 menetapkan klasifikasi status gizi dengan ambang batas pada tabel 2.1.

Tabel 2.4 Klasifikasi Status Gizi Kementerian Kesehatan RI[12]

Indeks	Kategori Status Gizi	Ambang Batas
Berat Badan Menurut Umur (BB/U)	Berat Badan Sangat Kurang(Severely Wasted)	< - 3 SD
	Berat Badan Kurang (Underweight)	-3 SD sd <-2 SD
	Berat Badan Normal	-2 SD sd +1 SD
	Resiko Berat Badan Lebih	>-1 SD
Tinggi Badan Menurut Umur (TB/U)	Sangat Pendek (Severely Stunted)	<-3 SD
	Pendek (Stunted)	-3 SD sd <-2 SD
	Normal	-2 SD sd +1 SD
	Tinggi	>+3 SD
Berat Badan menurut Panjang atau Tinggi Badan (BB/TB)	Gizi Buruk	<-3 SD
	Gizi Kurang	-3 SD sd <-2 SD
	Gizi Normal	-2 SD sd +1 SD
	Beresiko Gizi Lebih	>+3 SD

Dari tabel 2.4 klasifikasi status gizi balita menurut kementerian kesehatan, diketahui bahwa indikator berat badan menurut umur (BB/U) menghasilkan status gizi balita dengan kondisi berat badan sangat kurang, berat badan kurang, berat badan normal dan resiko berat badan lebih. Untuk indikator Tinggi badan menurut Umur (TB/U) menghasilkan klasifikasi status gizi balita sangat pendek, pendek, normal dan tinggi. Sedangkan indikator berat badan menurut tinggi badan (BB/TB) menghasilkan status gizi, gizi buruk, gizi kurang, gizi normal dan beresiko gizi lebih [12].

Nilai ambang batas digunakan sebagai batas nilai pada tabel baku standar berat badan/tinggi badan. Apabila diketahui berat badan balita berada pada simpang baku -1 maka masih dalam kategori normal jika dilihat dari nilai tengah

atau median. Begitupula dengan simpang baku +1. Namun jika beratbadan/tinggi badan berada pada -2 atau -3 maka ada 2 indikasi. Dimana kondisi dapat berupa gizi kurang jika dilihat berdasarkan indikator berat badan, atau pendek (stunted) apabila dilihat berdasarkan indikator tinggi badan. Sedangkan untuk nilai berat badan/tinggi badan berada pada simpang baku +2 atau +3, maka perlu diwaspadai status gizi berada pada resiko berlebih atau obesitas.

2.2.3 Variable Pengukuran Status Gizi

Ada beberapa cara melakukan penilaian status gizi balita pada kelompok masyarakat. Salah satu yang sering digunakan dalam posyandu adalah pengukuran tubuh manusia atau antropometri[13]. Dalam pengukuran antropometri terdapat 4 variabel pengukuran, dari 4 variabel ini dapat dilakukan penentuan status gizi berdasarkan indikator berat badan maupun tinggi badan. Variabel tersebut adalah sebagai berikut :

a. Jenis Kelamin

Menurut KBBI Jenis kelamin diartikan sebagai sifat (keadaan) laki-laki atau perempuan seseorang. Jenis kelamin akan mempengaruhi pertumbuhan tubuh manusia. Laki-laki mempunyai potensi berat dan tinggi badan lebih dibandingkan perempuan. Bayi baru lahir dengan jenis kelamin laki-laki mempunyai berat badan optimal 3,3 kg dan panjang 49,9 cm, sedangkan bayi perempuan berat badan optimalnya 3,2 kg dan panjang bayi 49,1 cm. Potensi tinggi badan optimal laki-laki dewasa (setelah usia 19 tahun) adalah 176,5 cm, sedangkan potensi tinggi badan perempuan dewasa 163,2 cm[6].

b. Umur

Umur merupakan lama waktu hidup seseorang durasi atau lama hidup seseorang dari saat lahir. Terdapat dua periode pertumbuhan cepat, yaitu pada usia bawah lima tahun (balita) dan periode remaja. Pada periode balita terjadi pertumbuhan semua jaringan tubuh terutama otak, dan balita lebih mudah mengalami sakit dan menderita kurang gizi[6]. Periode balita merupakan dasar pembentukan kepribadian anak, sehingga memerlukan perhatian khusus. Berdasarkan Standar Pemantauan Pertumbuhan, umur ditetapkan sebagai bulan penuh (30 hari). Sebagai contoh umur 23 hari = 0 bulan, umur 3 bulan 14 hari

= 3 bulan, umur 3 bulan 29 hari = 3 bulan [13]. Untuk keperluan penilaian status gizi maka umur dinyatakan dalam satuan bulan penuh.

c. Berat Badan

Berat badan menggambarkan jumlah protein, lemak, air, dan mineral yang terdapat di dalam tubuh. Terdapat beberapa alasan kenapa berat badan digunakan sebagai salah satu variabel pengukuran status gizi. Alasan tersebut di antaranya adalah berat badan sangat peka terhadap perubahan yang mendadak baik karena infeksi ataupun konsumsi makanan yang menurun. Berat badan ini dinyatakan dalam indikator BB/U. Berat badan paling mudah digunakan karena hanya memerlukan satu pengukuran. Tetapi kurang dapat menggambarkan kecenderungan perubahan status gizi dari waktu yang lama.

d. Tinggi Badan

Panjang badan atau tinggi badan merupakan parameter antropometri untuk pertumbuhan linier. Tinggi badan merupakan parameter antropometri untuk menilai pertumbuhan panjang atau tinggi badan. Dalam hal ini tinggi badan memberikan gambaran fungsi pertumbuhan yang dilihat dari keadaan kurus kering dan kecil pendek. Tinggi badan sangat baik untuk melihat keadaan gizi masa lalu. Tinggi badan juga digunakan dalam indikator penentu status gizi berdasarkan umur (TB/U) maupun status gizi berdasarkan (BB/TB)[13].

2.3 Pembelajaran Mesin

Saat ini pembelajaran mesin (*Machine Learning*) dan ilmu data semakin populer. Bidang ini diperkirakan akan tumbuh secara eksponensial di tahun-tahun mendatang. Secara ilmiah pembelajaran mesin mengacu pada pengamatan data yang dikumpulkan dari fenomena yang sedang dipelajari. Contohnya dalam bidang manufaktur, data menangkap informasi tentang kinerja mesin, hingga tingkat produksi, dari data-data yang dikumpulkan ini dapat diprediksi peningkatan tingkat produksi dengan memperhitungkan komponen-komponen penunjangnya. Adanya pembelajaran mesin ini dapat dijadikan sebagai teknologi yang memungkinkan penyimpanan dan pemrosesan data yang efektif dengan biaya yang rendah. Pada dasarnya pembelajaran mesin (*Machine learning*) adalah cabang ilmu komputer yang memanfaatkan pengalaman masa lalu untuk belajar

dan menggunakan pengetahuannya untuk membuat keputusan di masa depan[14]. Tujuan dari pembelajaran mesin adalah untuk menggeneralisasikan pola yang dapat dideteksi atau membuat sebuah modeling dari data yang diberikan[14, p. 32]. Pembelajaran mesin secara luas diklasifikasikan kedalam dua kategori :

a. Supervised Learning

Pembelajaran mesin merupakan proses pembelajaran mesin yang pembelajarannya di bawah pengawasan[14]. Dalam artian terdapat data yang memiliki label untuk membangun sebuah model. Dalam supervised learning terdapat variabel x dan variable output yang biasa disebut variabel y . Tujuan dari supervised learning ini untuk mempelajari fungsi pemetaan dari variabel x ke variabel y . Supervised learning dapat dimanfaatkan untuk klasifikasi atau regresi.

b. Unsupervised learning

Berbeda dengan supervised learning, unsupervised learning merupakan proses pembelajaran yang lebih bebas dan tidak ada pengawasan[14]. Secara bebas berarti dalam proses eksplorasi data tidak memiliki label, sehingga model akan ditentukan dengan menemukan struktur dan pola didalam dataset secara langsung. Kategori pembelajaran yang termasuk dalam unsupervised learning adalah clustering.

Pengembangan dan penerapan model pembelajaran mesin melibatkan beberapa langkah-langkah yang serupa dengan proses pemodelan statistik[15]. Proses pemodelan mesin ini digunakan untuk mengembangkan, memvalidasi dan mengimplementasikan sebuah data sehingga menjadi sebuah pola tertentu, yang salah satunya dapat digunakan untuk pembuat keputusan[15]. Langkah-langkahnya adalah sebagai berikut :

a. Pengumpulan Data

Data merupakan hal yang sangat penting dalam pembelajaran mesin. Data menjadi kunci pengambilan keputusan dari hasil informasi yang didapatkan. Data pembelajaran mesin dikumpulkan langsung dari sumber terstruktur dan tidak terstruktur seperti suara, gambar atau teks.

b. Persiapan Data

Pada langkah ini, data mentah diolah untuk mempermudah pembelajaran. Proses ini disebut dengan preprocessing. Preprocessing mencakup serangkaian proses pada data yang diamati, seperti noise pada data, pemfilteran, penghapusan outlier dan pengolahan data yang hilang[15].

c. Analisa Data dan Rekayasa Fitur

Data perlu dianalisis secara berurutan untuk menemukan pola atau hubungan antar variabel. Dari pola yang terbentuk dijadikan sebagai model yang nantinya dapat dijadikan bahan untuk klasifikasi, atau clustering sehingga hasilnya dapat menjadi informasi yang digunakan untuk pengambilan keputusan[14].

d. Train dan Data Validasi

Proses pembelajaran mesin (Machine Learning) dalam hal ini mesin yang dikembangkan untuk bisa belajar dengan sendiri dengan data, maka perlu ada proses “belajar”[14]. Proses “belajar” yang dilakukan pada pembelajaran mesin (machine learning) yaitu dengan pembagian dataset kedalam data training. Data training sendiri merupakan kumpulan data dari dataset yang digunakan untuk melatih algoritma. Dari hasil pembelajaran algoritma dengan data training, maka perlu adanya pembuktian apakah model yang terbentuk telah sesuai dalam artian akurat di validasi dan akurat juga di kenyataannya dilakukanlah testing. Testing dalam hal ini menggunakan data test yang berupa data baru dengan karakteristik yang sama dengan data train, sehingga algoritma dapat mengenali dan menghasilkan informasi berdasarkan data.

e. Pengujian Model

Setelah model terbentuk dari data, kinerja dari model ini akan diperiksa terhadap data uji. Dari pengujian model juga didapatkan apakah sebuah informasi cukup signifikan dan valid.

f. Penerapan Pembelajaran

Dari hasil modeling dan pengujian, maka diketahui sebuah model pembelajaran ini dapat diterapkan dalam sebuah sisten. Sebagai contoh dari pembelajaran bisa menjadi sisten rekomendasi, pengambilan keputusan, atau klasifikasi dari sekumpulan data.

2.4 Naïve Bayes

Naïve Bayes atau aturan bayes adalah metode penggalian data atau pembelajaran mesin yang digunakan untuk membuat model dengan pengklasifikasian probabilistik sederhana dengan menjumlahkan kombinasi nilai dari dataset[16]. Algoritma menggunakan teorema Bayes mengasumsikan semua atribut independen atau tidak saling ketergantungan pada nilai variable. Definisi lain mengatakan Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik dengan memprediksi peluang di masa depan berdasarkan pengalaman sebelumnya.

2.4.1 Persamaan Metode *Naïve Bayes*

Persamaan teorema Bayes [16]:

$$P(H | X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2.1)$$

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan class spesifik

P(H|X): Probabilitas hipotesis H berdasarkan Kondisi X (Probabilitas posterior)

P(X|H): Probabilitas berdasarkan kondisi pada hipotesis

P(H) : Probabilitas hipotesis H (probabilitas prior)

P(X) : Probabilitas X

Pengklasifikasian *Naïve Bayes* digunakan untuk menentukan kelas yang cocok untuk sample yang dianalisis. Maka teorema bayes disesuaikan pada persamaan (2.2).

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \quad (2.2)$$

Variable C menunjukan kelas, sementara variable F1.....Fn sebagai atribut karakteristik yang digunakan untuk melakukan klasifikasi. Sehingga rumus tersebut menjelaskan peluang masuknya sample dengan karakteristik tertentu pada kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel) dikali dengan peluang karakteristik yang muncul sampel pada kelas C, dan dibagi dengan peluang karakteristik yang muncul dari keseluruhan sampel[16].

Untuk data kontinyu maka perhitungan klasifikasi menggunakan rumus Densitas Gauss [15] seperti pada persamaan (2.3):

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x_i - \mu)^2}{2\sigma^2}} \quad (2.3)$$

Keterangan :

- P : Peluang
- X_i : Atribut ke i
- X_i : Nilai atribut i
- Y : Kelas yang dicari
- y_i : Sub Kelas Y yang dicari
- μ : Mean, rata-rata seluruh atribut
- σ : Standar Deviasi dari seluruh atribut

Densitas Gauss digunakan untuk menghitung nilai maksimum probabilitas dari kelas klasifikasi *Naive Bayes*. Pada tahap pembacaan dataset data kontinyu akan dihitung nilai probabilitas, sedangkan data dengan nilai numerik akan dihitung mean dan standar deviasi.

2.5 Regresi Linier Berganda

Terdapat perbedaan yang mendasar antara analisis kolerasi dengan regresi, jika pada analisis kolerasi digunakan untuk mencari arah dan kuatnya hubungan antara dua variabel atau lebih, untuk analisis regresi digunakan untuk

memprediksi seberapa jauh perubahan nilai variabel dependen, apabila variabel independen di manipulasi[17]. Dalam hal ini variabel independen dapat dirubah atau dihilangkan. Pemanfaatan analisis regresi ini dilakukan untuk mengetahui apakah naik dan turunnya variabel dependen dapat dilakukan melalui peningkatan variabel independen. Dalam analisis redresi terdapat Regresi Linier sederhana yang dipergunakan untuk menggambarkan garis yang menunjukkan arah hubungan antar variabel, serta digunakan untuk melakukan prediksi respon berdasarkan variabel perdiktor tunggal[17]. Namun, dalam praktiknya sering ditemukan kasus klasifikasi atau clustering yang memiliki lebih dari satu prediktor. Jika kita menggunakan model regresi linier sederhana untuk setiap prediktor tentunya kurang efektif, maka dari itu perlu pendekatan yang lebih baik dengan memperluas model regresi linier sederhana sehingga dapat mengakomodasi beberapa prediktor sekaligus. Secara umum jika kita memiliki dua hingga lebih prediktor yang berbeda, maka menggunakan model regresi linier berganda. Analisis regresi linier berganda digunakan oleh peneliti, dengan persamaan regresi linier berganda[17] pada persamaan 2.4 berikut ini.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e \quad (2.4)$$

Keterangan :

Y : Variabel Dependen

B : Koefisien Determinasi

X : Variabel Independen (prediktor)

e : Error term

Persamaan 2.4 merupakan persamaan regresi untuk X prediktor. Untuk membuat analisis regresi maka data setiap variabel harus tersedia. Selanjutnya berdasarkan data itu peneliti harus menemukan persamaan melalui perhitungan.

Susunan persamaan matriks untuk perhitungan koefisien determinasi (β) dapat dituliskan pada persamaan 2.5.

$$\begin{bmatrix} n & \Sigma X_1 & \Sigma X_2 & \Sigma X_3 & \Sigma X_4 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 \Sigma X_2 & \Sigma X_1 \Sigma X_3 & \Sigma X_1 \Sigma X_4 \\ \Sigma X_2 & \Sigma X_1 \Sigma X_2 & \Sigma X_2^2 & \Sigma X_2 \Sigma X_3 & \Sigma X_2 \Sigma X_4 \\ \Sigma X_3 & \Sigma X_1 \Sigma X_3 & \Sigma X_2 \Sigma X_3 & \Sigma X_3^2 & \Sigma X_3 \Sigma X_4 \\ \Sigma X_4 & \Sigma X_1 \Sigma X_4 & \Sigma X_2 \Sigma X_4 & \Sigma X_3 \Sigma X_4 & \Sigma X_4^2 \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \Sigma Y \\ \Sigma X_1 Y \\ \Sigma X_2 Y \\ \Sigma X_3 Y \\ \Sigma X_4 Y \end{bmatrix} \quad (2.5)$$

Maka koefisien determinasi (β) diperoleh dengan invers matrik (X^{-1}) dikali dengan matriks Y :

$$\rightarrow X \cdot \beta = Y$$

$$\beta = X^{-1} Y$$

$$\beta = \begin{bmatrix} n & \Sigma X_1 & \Sigma X_2 & \Sigma X_3 & \Sigma X_4 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 \Sigma X_2 & \Sigma X_1 \Sigma X_3 & \Sigma X_1 \Sigma X_4 \\ \Sigma X_2 & \Sigma X_1 \Sigma X_2 & \Sigma X_2^2 & \Sigma X_2 \Sigma X_3 & \Sigma X_2 \Sigma X_4 \\ \Sigma X_3 & \Sigma X_1 \Sigma X_3 & \Sigma X_2 \Sigma X_3 & \Sigma X_3^2 & \Sigma X_3 \Sigma X_4 \\ \Sigma X_4 & \Sigma X_1 \Sigma X_4 & \Sigma X_2 \Sigma X_4 & \Sigma X_3 \Sigma X_4 & \Sigma X_4^2 \end{bmatrix}^{-1} \times \begin{bmatrix} \Sigma Y \\ \Sigma X_1 Y \\ \Sigma X_2 Y \\ \Sigma X_3 Y \\ \Sigma X_4 Y \end{bmatrix} \quad (2.6)$$

Dari persamaan 2.6 diperoleh nilai masing-masing dari koefisien determinasi (β) yang kemudian nilai tersebut akan menjadi nilai koefisien regresi dari masing-masing variabel independen.

2.5.1 Uji Linearitas Regresi

Salah satu asumsi dari analisis regresi adalah linearitas. Dalam hal ini menunjukkan garis regresi antara X dan Y yang membentuk garis linear. Berikut ini persamaan-persamaan [17, p. 295] yang digunakan dalam uji linearitas :

Jumlah Kuadrat Total

$$JKT = \sum y_i^2 - \frac{\Sigma y_i^2}{n} \quad (2.7)$$

Jumlah Kuadrat Regresi

$$JKR = b_1 \left\{ \Sigma (x_{1i} y) - \frac{\Sigma (x_{1i}) \Sigma (y_i)}{n} \right\} + \dots + b_n \left\{ \Sigma (x_{pi} y) - \frac{\Sigma (x_{pi}) \Sigma (y_i)}{n} \right\} \quad (2.8)$$

Jumlah Kuadrat Galat

$$JKG = JKT - JKR \quad (2.9)$$

Koefisien Determinasi

$$R^2 = \frac{JKR}{JKT} \quad (2.10)$$

Nlai Dugaan

$$S^2 = \frac{R^2}{n-p} \quad (2.11)$$

Seperti yang dibahas pada bagian sebelumnya, regresi linier berganda digunakan untuk menghitung prediksi (Y) dan memeriksa variabel yang terkait dengan hasil, maka langkah selanjutnya dilakukan seleksi variabel untuk menentukan prediktor mana yang terkait dengan hasil prediksi. Pemilihan variabel dilakukan dengan mencoba model yang berbeda yang berisi subset yang berbeda dari prediktor. Misalnya jika memiliki dua prediktor, maka kita dapat langsung mempertimbangkan empat model antara lain : model yang tidak mengandung variabel, model yang hanya mengandung satu prediktor, atau model yang berisi semua variabel prediktor. Hal ini dapat dilakukan apabila prediktor memiliki jumlah yang sedikit, namun apabila prediktor memiliki jumlah yang lebih banyak tentunya tidak efektif. Diperlukan uji koefisien korelasi (R) yang digunakan untuk mengukur seberapa besar hubungan variabel bebas yang diteliti terhadap variabel terikat